Binational Integration of National Land Use/Land Cover Datasets in the United States-Mexico Border Region

U.S. Geological Survey
U.S.-Mexico Border Environmental Health Initiative
http://borderhealth.cr.usgs.gov

June 2006

by Zachary D. Wilson
USGS Texas Water Science Center
Austin, Texas

Introduction

An integral part of the U.S. Geological Survey's (USGS) U.S.-Mexico Border Environmental Health Initiative (BEHI) geospatial database is a binationally integrated land use/land cover (LULC) dataset. Among many other uses, LULC data can be used to analyze landscape change, to provide data for hydrologic modeling applications, to statistically analyze landscape fragmentation, and for display as a base layer for regional maps (Lillesand and Kiefer, 2000). The BEHI project has created a binationally integrated LULC dataset using existing LULC data from the United States and Mexico.

Both the United States and Mexico have programs to produce LULC datasets using a nationally consistent classification scheme. The USGS National Land Cover Dataset 1992 (NLCD92) and the National Land Cover Database 2001 (NLCD01) are based on the LULC classification system for remotely sensed data described by Anderson and others (1976). Both of these datasets are available in raster format at 30-meter resolution (Homer and others, 2004; USGS, 2000). In Mexico, the Instituto Nacional de Geografía, Estadística, e Informática (INEGI), the National Geography, Statistics, and Information Technology Institute, is responsible for mapping LULC. INEGI produces the 1:250,000-scale Mapa de Uso de Suelo y Vegetación (Land Use and Vegetation Map) based on visual interpretation of remotely sensed imagery using unified regional vegetation classification systems developed over the last 70 years by many scientists, including Leopold, Muller, and Rzedowski (INEGI, 1993). The Mexican LULC data are available in digital vector format.

Though each country's classification system is consistent within the country's own borders, the classes defined by the respective classification systems do not represent a one-to-one relationship across the border. Integration of the U.S. and Mexican data required the creation of a generalized (modified Anderson Level I) binational classification system to which both countries' LULC data could be reclassified (Anderson, 1976). The binational data integration described here includes the entire U.S.-Mexico Border region (see Figure 1) as defined by Woodward and Durall (1994) and adopted by the BEHI project (USGS, 2004a; 2004b).

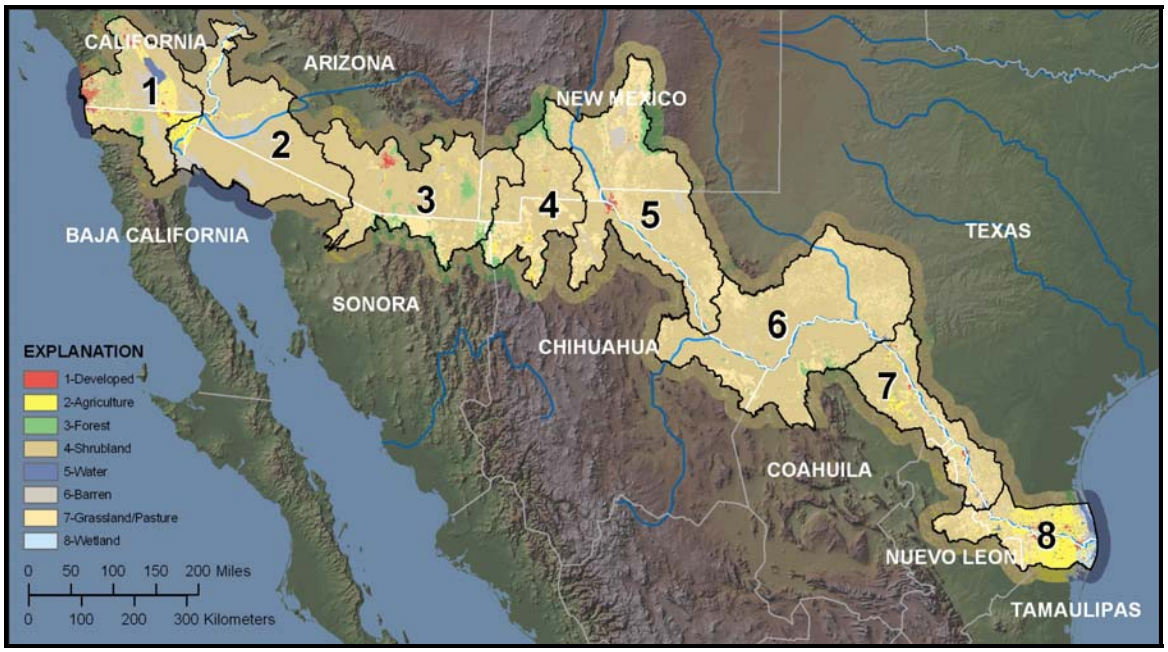

Figure 1. USGS U.S.-Mexico Border Environmental Health Initiative study area boundary with sub-areas and binationally-integrated land use/land cover.

Land Use/Land Cover Data

The two sources of LULC data are:

Mexico: INEGI 1993 Uso de Suelo Serie II 1:250,000-scale vector data

United States: USGS NLCD92 30m resolution raster data

The INEGI Uso de Suelo and the NLCD92 datasets were chosen because they represent a consistent nationwide classification system for their respective countries. Landsat images were the source data used to classify and create both the U.S. and Mexican LULC datasets. However, the classifications were conducted using different methods. USGS, in cooperation with the U.S. Environmental Protection Agency (EPA), created the NLCD92 using unsupervised classification methods along with ancillary datasets, such as elevation data, local land cover datasets, and wetlands inventories. The final dataset is a 21-class LULC raster dataset with 30-meter spatial resolution. INEGI's LULC dataset was also derived from Landsat imagery, but was created utilizing manual methods. LULC types were visually interpreted from the Landsat imagery. Polygons were digitized to delineate LULC types and then verified with fieldwork. The final INEGI dataset consists of a polygon layer, which has more than 100 LULC classes in the BEHI study area.

Binational Classification Scheme

The creation of a land cover dataset requires a classification scheme that suits the purpose for which the dataset is being created. For example, a forest scientist may need more forest classes defined than other classes, while an agricultural scientist may be more interested in agricultural classes and only need one forest class. Constraints on time and funding can also dictate whether

certain classes will be included or not. Differences in the design of classification systems affect the ways in which a land cover dataset may be used. It is quite easy to take the forest scientist's land cover dataset and aggregate the many forest classes into one. However, it is not possible to create many forest classes from the agricultural scientist's single forest class without referring to the original imagery used in creating the dataset. Each of these example datasets were created independently by scientists pursuing different ends.

Differences between the U.S. and Mexican LULC datasets are similar to those of the hypothetical datasets described above. For example, the Mexican LULC dataset has only one class for urban areas, while the U.S. dataset has several. Conversely, the U.S. dataset has only one class for shrubland, while the Mexican dataset has several. In terms of the classification scheme described by Anderson and others, the U.S. LULC is an Anderson Level II, while the Mexican dataset is mostly an Anderson Level III (1976). The Mexican dataset contains more than five times the number of classes as the U.S. dataset, with LULC classes describing cover types down to the dominant species level in some cases. Binationally integrating the U.S. and Mexican LULC datasets required the creation of a binational scheme, independent of the two datasets. One example of a binational class is "forest". All forest classes in the United States, including deciduous, evergreen, and mixed, were reclassified to the binational forest class. The Mexican forest classes, for example pine forest, oak forest, and cedar forest, were also reclassified to the binational forest class. The reclassification causes the loss of some of the existing detail in the original data, but it allows the integration of dissimilar classes. The Mexican data refer to forests by the composition of specific species, while the U.S. data refer to forest type. One feasible way to integrate the two classification schemes is to create a binational scheme that accommodates both datasets.

The binational scheme was created by individually analyzing all original LULC classes and determining how they could be reclassified to fit into a binational classification scheme, a slightly modified Anderson Level I classification (see Tables 1–4). Anderson and others suggest the following LULC classes for a Level I classification: urban or built-up land, agricultural land, rangeland, forestland, water, wetland, barren land, tundra, and perennial ice or snow. In the modified Anderson Level I binational classification scheme, it is possible to split the Anderson Level I rangeland class into two classes, shrubland and grassland/pasture. The remaining classes are identical in both the Anderson Level I classification scheme and the modified classification scheme. Tundra and perennial ice were excluded because they did not exist in the study area.

Anderson Level I	Modified Level I
urban or built-up land	urban or built-up land
agricultural land	agricultural land
rangeland	grassland
forestland	shrubland
water	forest
wetland	water
barren land	wetland
	barren land

Not surprisingly, one of the LULC classes did not fit perfectly into the binational schema. The Mexico LULC dataset contains a class designated as "halophilic vegetation". It occurs along the Gulf Coast as well as inland in arid landscapes. In order to create a better fit for binational integration, the halophilic vegetation class polygons that intersected the coast were reclassified as wetland. The remaining halophilic vegetation polygons were converted to shrubland. The

decision to reclassify was based on evidence from the wetland class in the US LULC dataset and satellite and aerial imagery.

Binational Integration

The process for creation of the binational LULC dataset consists of the following basic geographic information system (GIS) procedures:

1. Convert input LULC datasets to the same format and coordinate system.
2. Reclassify LULC datasets to a modified Anderson Level I.
3. Clip USA and Mexico LULC datasets using a common international boundary.
4. Merge clipped and reclassified data into binational LULC dataset.

The full integration process consists of several pre-processing steps, a geoprocessing model created with ESRI ArcGIS 9.0 software, and some post-processing steps. The detailed process steps are as follows:

1. Create a common unofficial international boundary by digitizing the center of the Rio Grande/Río Bravo using 1-meter resolution orthoimagery from 2004.
2. Combine the newly digitized portion of the international boundary that follows the Rio Grande/Río Bravo with the existing boundary along the western half of the border.
3. Use the common international boundary to create polygon masks for USA and Mexico for clipping land cover data.
4. Convert polygon masks to 30-meter resolution raster data.
5. Select “halophilic vegetation” polygons that intersect the coastline and reclassify them to “wetland”.
6. Convert Mexico LULC vector land cover data to 30-meter resolution raster data.
7. Reclassify USA LULC raster data to modified Anderson Level I using reclassification table (see Table 2).
8. Reclassify Mexico LULC raster data to modified Anderson Level I using reclassification table (see Tables 3 and 4).
9. Clip reclassified USA and Mexico LULC raster data using raster masks.
10. Mosaic clipped reclassified USA and Mexico LULC raster data to a new binational LULC 30-meter raster dataset (see Table 1).
11. Clip binational LULC mosaic to study area boundary.
12. Assign display colors for each class using a color map.

Limitations

While the binational classification scheme is robust and the source imagery for the classification of LULC are the same, the method of classification of the imagery and the format of the original datasets pose a potential problem. As mentioned earlier, the Mexican data were processed using manual digitization methods. Image interpreters digitized polygons based on what they saw in the image and the final product was verified with field work. The U.S. LULC data were classified using automated techniques that resulted in a raster dataset. The differences in format of the original LULC data make it difficult to analyze the integrated binational dataset. LULC represented by polygons tends to present a more homogeneous picture of the landscape, while raster data are able to represent more heterogeneity. The same landscape would be represented differently if classified using the two distinct methods employed by the U.S. and Mexico. Caution

should be exercised when analyzing the final integrated binational LULC dataset. However, the binational LULC dataset does provide a good qualitative representation of regional patterns in LULC.

Summary

One of the most important base datasets in any geospatial database is LULC. To that end, the BEHI project has produced a LULC dataset, in line with its goal of producing a binational geospatial database along the U.S-Mexico Border. The result of the reclassification methodology describe above is a binationally integrated 30-meter resolution LULC raster dataset. This dataset represents the integration of "best-available" national LULC data and has limitations. It also represents one approach (integration of existing data) of many possible approaches (for example, classification of satellite imagery using existing data as training data). The dataset is intended for use at broad spatial scales, i.e. from about 1:1,000,000 to the entire U.S.-Mexico border, and should be considered inappropriate for local scales.

Tables

Table 1. Binational LULC classes with description

Binational LULC class	Binational LULC description
1	Developed
2	Agriculture
3	Forest
4	Shrubland
5	Water
6	Barren
7	Grassland/Pasture
8	Wetland

Table 2. NLCD92 classes with description and binational LULC class

NLCD92 class	NLCD92 description	Binational LULC class
11	Open Water	5
12	Perennial Ice/Snow	6
21	Low Intensity Residential	1
22	High Intensity Residential	1
23	Commercial, Industrial, Transportation	1
31	Bare Rock, Sand, Clay	6
32	Quarries, Strip Mines, Gravel Pits	6
33	Transitional	6
41	Deciduous Forest	3

NLCD92 class	NLCD92 description	Binational LULC class
42	Evergreen Forest	3
43	Mixed Forest	3
51	Shrubland	4
61	Orchards, Vineyards, Other	2
71	Grasslands, Herbaceous	7
81	Pasture, Hay	7
82	Row Crops	2
83	Small Grains	2
84	Fallow	2
85	Urban, Recreational Grasses	1
91	Woody Wetlands	8
92	Emergent Herbaceous Wetlands	8

Table 3. Mexico LULC classes and binational equivalents, sorted by FC code. FC code values consist of a two or three digit code, sometimes prefixed by an “E” or followed by another code. The “E” prefix denotes areas subjected to significant erosion. If the FC code is followed by another code, the second code is a secondary vegetation type. The first letter of each code is its general class (B = forest, M = shrubland, P = grassland, R = agriculture, T = temporary agriculture, V = various types of vegetation, ZU = urban)

FC	CLAVEFOT	Binational LULC class
6122	HA	2
6152	RA	2
6154	RAS	2
6156	RAP	2
6162	RS	2
6164	RSA	2
6172	RP	2
6174	RPA	2
6182	ReA	2
6211	E-[R]	2
6212	[R]	2
6213	E-TA	2
6214	TA	2
6218	TAP	2
6234	TP	2
6250	DV	6
6251	E-DV	6
6282	BS	3

FC	CLAVEFOT	Binational LULC class
6284	BS/VSa	3
6292	H2O	5
6298	BQ	3
6300	BQ/VSa	3
6302	BQ/VSA	3
6306	BQP	3
6308	BQP/VSa	3
6314	BG	3
6322	BA	3
6324	BA/VSa	3
6330	BP	3
6332	BP/VSa	3
6338	BPQ	3
6340	BPQ/VSa	3
6346	BJ	3
6404	MC/MB	4
6414	MC/MN	4
6429	E-MDM/MI	4
6430	MDM/MI	4
6431	E-MDM/MB	4
6432	MDM/MB	4
6433	E-MDM/ME	4
6434	MDM/ME	4
6440	MDM/MCH	4
6453	E-MDM/VSa	4
6454	MDM/VSa	4
6458	MDR/MI	4
6459	E-MDR/MB	4
6460	MDR/MB	4
6462	MDR/ME	4
6473	E-MDR/MR	4
6474	MDR/MR	4
6482	MDR/VSa	4
6486	MET/MI	4
6487	E-MET/MB	4
6488	MET/MB	4
6489	E-MET/ME	4
6490	MET/ME	4
6509	E-MET/VSa	4

FC	CLAVEFOT	Binational LULC class
6510	MET/VSa	4
6525	ZU	1
6626	MSM/MI	4
6627	E-MSM/MB	4
6628	MSM/MB	4
6649	E-MSM/VSa	4
6650	MSM/VSa	4
6692	ML	4
6694	ML/VSa	4
6697	E-MU	4
6721	E-MK	4
6722	MK	4
6725	E-MK/VSa	4
6726	MK/VSa	4
6730	VP	4
6746	VT	4
6754	VU	4
6760	VG	4
6771	E-VH	4*
6772	VH	4*
6774	VH/VSa	4*
6781	E-PC	7
6782	PC	7
6784	PY	7
6786	PY/VSa	7
6790	PH	7
6795	E-PI	7
6796	PI	7
6797	E-PN	7
6798	PN	7
6800	PN/VSa	7
6810	VW	7
6999	n/a	8

*along the coast, features in this class were assigned to the 8 – Wetland class, and given a modified FC code of 6999.

Table 4. Mexico LULC classes and binational equivalents, sorted by CLAVEFOT code.

FC	CLAVEFOT	Binational LULC class

FC	CLAVEFOT	Binational LULC class
6212	[R]	2
6322	BA	3
6324	BA/VSa	3
6314	BG	3
6346	BJ	3
6330	BP	3
6332	BP/VSa	3
6338	BPQ	3
6340	BPQ/VSa	3
6298	BQ	3
6302	BQ/VSA	3
6300	BQ/VSa	3
6306	BQP	3
6308	BQP/VSa	3
6282	BS	3
6284	BS/VSa	3
6250	DV	6
6211	E-[R]	2
6251	E-DV	6
6431	E-MDM/MB	4
6433	E-MDM/ME	4
6429	E-MDM/MI	4
6453	E-MDM/VSa	4
6459	E-MDR/MB	4
6473	E-MDR/MR	4
6487	E-MET/MB	4
6489	E-MET/ME	4
6509	E-MET/VSa	4
6721	E-MK	4
6725	E-MK/VSa	4
6627	E-MSM/MB	4
6649	E-MSM/VSa	4
6697	E-MU	4
6781	E-PC	7
6795	E-PI	7
6797	E-PN	7
6213	E-TA	2
6771	E-VH	4*
6292	H2O	5

FC	CLAVEFOT	Binational LULC class
6122	HA	2
6404	MC/MB	4
6414	MC/MN	4
6432	MDM/MB	4
6440	MDM/MCH	4
6434	MDM/ME	4
6430	MDM/MI	4
6454	MDM/VSa	4
6460	MDR/MB	4
6462	MDR/ME	4
6458	MDR/MI	4
6474	MDR/MR	4
6482	MDR/VSa	4
6488	MET/MB	4
6490	MET/ME	4
6486	MET/MI	4
6510	MET/VSa	4
6722	MK	4
6726	MK/VSa	4
6692	ML	4
6694	ML/VSa	4
6628	MSM/MB	4
6626	MSM/MI	4
6650	MSM/VSa	4
6782	PC	7
6790	PH	7
6796	PI	7
6798	PN	7
6800	PN/VSa	7
6784	PY	7
6786	PY/VSa	7
6152	RA	2
6156	RAP	2
6154	RAS	2
6182	ReA	2
6172	RP	2
6174	RPA	2
6162	RS	2
6164	RSA	2

FC	CLAVEFOT	Binational LULC class
6214	TA	2
6218	TAP	2
6234	TP	2
6760	VG	4
6772	VH	4*
6774	VH/VSa	4*
6730	VP	4
6746	VT	4
6754	VU	4
6810	VW	7
6525	ZU	1
6999	n/a	8

*along the coast, features in this class were assigned to the 8 – Wetland class, and given a modified FC code of 6999.

References

Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer, 1976, A Land Use and Land Cover Classification System for Use with Remote Sensor Data: U.S. Geological Survey Professional Paper 964. Last accessed June 2, 2006 at: http://landcover.usgs.gov/pdf/anderson.pdf

Homer, C., C. Huang, L. Yang, B. Wylie, and M. Coan, 2004, Development of a 2001 National Land-Cover Database for the United States: Photogrammetric Engineering & Remote Sensing 70(7): 829–840. Last accessed June 2, 2006 at: http://www.mrlc.gov/pdfs/July_PERS.pdf

Instituto Nacional de Estadística, Informática e Geografía (INEGI), 1993, Guía Para la Interpretación de Información Cartográfica Impresa y Digital de Uso de Suelo.

Lillesand, T.M and R.W. Kiefer, 2000, Remote Sensing and Image Interpretation, Fourth Edition: John Wiley & Sons, Inc., New York.

U.S. Geological Survey, 2000, National Land Cover Dataset: U.S. Geological Survey Fact Sheet 108-00. Last accessed June 2, 2006 at: http://erg.usgs.gov/isb/pubs/factsheets/fs10800.html

U.S. Geological Survey, 2004a, Science Data in Support of Environmental Health Studies in the U.S.-Mexico Border Region: U.S. Geological Survey Fact Sheet 2004–3013.

U.S. Geological Survey, 2004b, Internet Map Service for Environmental Health in the U.S.-Mexico Border Region: U.S. Geological Survey Fact Sheet 2004-3140.

Woodward, D.G. and R.A. Durall, 1996, United States-Mexico Border Area, as Delineated by a Shared-Water Resources Perspective: U.S. Department of the Interior Field Coordinating Committee Fact Sheet 1.

Links

USGS U.S.-Mexico Border Environmental Health Initiative
http://borderhealth.cr.usgs.gov

Instituto Nacional de Estadística, Geografía, e Informática (INEGI)
http://www.inegi.gob.mx

USGS National Land Cover Dataset (NLCD)
http://landcover.usgs.gov